

A COMPARATIVE STUDY OF AI MODELS FOR LUNG CANCER PREDICTION AND CLINICAL DECISION SUPPORT

A. Sirajdeen¹, S. Rajakumaran^{2*}

1. Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalai Nagar, Chidambaram, Tamil Nadu, India
2. Assistant Professor, Department of Computer Science, Government Arts and Science College, Mettur, Salem, Tamil Nadu, India

Abstract

Artificial-intelligence (AI) research on lung cancer in 2026 shows rapid growth across lightweight convolutional networks, CNN–transformer hybrids, capsule-inspired models, segmentation-driven pipelines, radiomics, multimodal fusion and reader-assistance systems. Model selection is now more than accuracy maximisation: a high closed-set score does not automatically make a model the best choice for low-resource deployment, indeterminate-nodule risk estimation, radiologist assistance or clinically interpretable decision support. This study comparatively analyses 20 peer-reviewed studies published in 2026, examining reported predictive performance together with computational efficiency, explainability, imaging modality, task definition and direct clinical comparison. Public-dataset classifiers frequently report accuracy above 96%, with several attention-based or compact architectures reaching 98–100% on selected hold-out sets. Clinically anchored studies present different evidence: a radiologist reader study increased cancer-detection AUC from 0.652 to 0.761 with AI support and reduced interpretation time by 12.9%; PET/CT malignancy modelling achieved an AUC of 0.78, non-inferior to the Herder model (AUC 0.73); and CT radiomics reached an AUC of 0.941, statistically comparable to expert radiologists. Multimodal fusion also produced clinically useful risk estimates, including a test AUC of 0.823 for stacked clinical–radiomics–deep-learning fusion. The comparative evidence supports task-specific model selection: compact networks for constrained hardware, attention hybrids for high-dimensional image representation, radiomics or fusion for risk stratification, and reader-validated systems for workflow integration.

Keywords

Lung Cancer, Computed Tomography, Deep Learning, Radiomics, Explainable AI, Pulmonary Nodules, Clinical Decision Support

1. Introduction

Lung cancer prediction has moved from a narrow architecture race towards a broader clinical engineering problem. Recent systems address several distinct tasks: image-level cancer classification, nodule malignancy estimation, lesion segmentation, histological subtype recognition, invasive-potential assessment, reader assistance and treatment-oriented risk stratification. A single performance number cannot represent all these objectives. Accuracy

is meaningful for balanced multiclass benchmarks, the area under the curve (AUC) is more informative for continuous malignancy risk, Dice or IoU better describes segmentation, and reader-study outcomes provide direct evidence of workflow benefit.

A second change concerns model design. Convolutional neural networks (CNNs) remain competitive, but transformer attention, hybrid feature fusion, capsule-inspired representations, radiomics and multimodal probability stacking now appear frequently in 2026 publications. Several models report near-perfect scores on curated public datasets [1]–

*Corresponding author: sirajdeen.89@gmail.com

[8], [18]. In contrast, clinically anchored investigations often report lower numerical discrimination while testing harder questions, such as performance against expert radiologists, comparison with the Brock or Herder models, or prediction under heterogeneous acquisition conditions [11]–[17]. Direct ranking across such settings can therefore reward dataset simplicity rather than practical usefulness.

The central problem addressed in this comparative study is model selection rather than score maximisation. A high-capacity hybrid model may offer excellent closed-set discrimination but at a substantial computational cost; a compact model may sacrifice a small amount of accuracy while reducing inference cost; a radiomics model may provide transparent feature-level evidence; and a multimodal model may improve risk stratification by combining morphology with clinical or serological information. A reader-assistance model may provide the strongest clinical relevance even when its headline AUC remains below that of the best public-dataset classifier.

The objective of this study is to provide a deployment-oriented synthesis of peer-reviewed lung-cancer AI studies published in 2026. The comparison covers the diagnostic task, imaging modality, architecture family, reported performance, explainability, efficiency and degree of clinical anchoring. All numerical results are author-reported values from the cited studies, and cross-paper values are not treated as a common leaderboard.

2. Comparative Study Design

Scope and evidence selection: Priority was given to studies on CT or PET/CT lung-cancer detection, pulmonary-nodule characterisation, lung-cancer classification, radiomics, multimodal fusion and AI-assisted reader evaluation. Preprints were excluded. Publisher pages, version-of-record records, PubMed entries and DOI metadata were used to verify references. Twenty studies satisfied the final comparative scope [1]–[20].

Comparison axes: Five axes structure the analysis. (i) Predictive performance covers accuracy, F1-score, AUC, sensitivity, specificity or segmentation overlap as reported by each publication. (ii)

Computational suitability covers model compactness, FLOPs, memory demand, inference time or explicit lightweight design. (iii) Interpretability covers Grad-CAM, attention visualisation, radiomic-feature analysis, SHAP or transparent risk variables. (iv) Clinical anchoring covers histopathology, guideline comparators, multi-centre data, reader studies and patient-management analysis. (v) Task alignment separates closed-set classification from continuous malignancy prediction and segmentation.

Interpretation rule: Cross-study differences are interpreted qualitatively rather than statistically because datasets, class definitions, partitions, imaging protocols and endpoints vary. Within-study comparisons receive greater weight. A model with a lower absolute AUC may carry stronger clinical evidence when evaluation includes difficult nodules, reader comparison, independent testing or guideline-based management thresholds.

3. Model Landscape

3.1 Lightweight and Efficient Networks

MiniConvNet targets resource-efficient non-small cell lung cancer (NSCLC) classification and reports 96% CT test accuracy, with five-fold accuracy near 96.6% [3]; its compact design directly addresses memory and inference constraints. A capsule-inspired EfficientNetB0 framework removes dynamic routing and the decoder, reducing reported complexity to 0.01 GFLOPs and approximately 4.85 million trainable parameters [5]. A perfect hold-out score is reported for its main IQ-OTH/NCCD experiment, while ten-fold mean accuracy drops to 92.61%, an important reminder that computational efficiency and hold-out performance require separate interpretation. A VGG16–U-Net++ ensemble reports 96% accuracy with a 3.1 GB GPU memory demand, positioning segmentation-assisted 2D processing as another resource-aware direction [19].

3.2 Attention, Transformer and Hybrid Models

CNN–transformer integration remains prominent. C-Swin combines VGG16 local features with shifted-window attention and reports 96.26% accuracy on IQ-OTH/NCCD [1]. CNN–HiFuse

combines multi-scale features with channel and spatial attention and reaches 98.12% held-out accuracy with $97.86 \pm 0.19\%$ five-fold accuracy [2]. LungDxFormer reports 97.35% accuracy on LIDC-IDRI and includes repeated-run significance testing [4]. A parallel spatial cross-attention ensemble reports 99.20% accuracy on LUNA16 [18]. A broad benchmark across CNN and transformer families reports a matching top F1-score of 0.9816 for strong CNN and transformer architectures, suggesting that no architecture family is universally superior [8].

3.3 Segmentation-Driven and Representation-Focused Systems

Segmentation remains valuable when classification depends on subtle nodule boundaries. A 3D Trans-DenseUNet++ pipeline couples volumetric segmentation with adaptive DenseNet–LSTM classification and reports 94.9% accuracy [6]. A dual-branch image-and-mask approach reports 98.04% accuracy in a 5,000-sample experiment, with lower results in a larger and more heterogeneous experiment [7]. A separate image-enhancement and

nodule-delineation pipeline reports a mean overlap of 0.83, sensitivity of 0.92 and 1.5 false positives per scan on 1,000 LIDC-IDRI scans [20]. These results show the complementary role of localisation quality, especially for downstream decision support.

3.4 Explainability and Feature-Guided Learning

Explainability appears in both deep and hand-crafted pipelines. ARXAF-Net combines active reinforcement learning, feature engineering, attention fusion and Grad-CAM, reporting 99% testing accuracy on a large curated CT collection [9]. Explainable convolutional analysis by Di Giammarco et al. emphasises visual explanation for lung adenocarcinoma recognition [10]. Radiomics offers a different route: selected intensity, shape and texture features support clinically legible reasoning without relying only on post-hoc heatmaps. Direct comparison with radiologists is therefore especially valuable for judging whether quantitative features capture diagnostically meaningful structure. The representative studies are summarised in Table 1.

Table 1. Representative 2026 Lung-Cancer AI Studies and Reported Evidence

<i>Study</i>	<i>Task / modality</i>	<i>Method family</i>	<i>Results</i>	<i>Main comparative value</i>
C-Swin [1]	CT, 3-class	CNN + Swin	Acc. 96.26%; F1 97.42%	Local/global feature fusion
CNN-HiFuse [2]	CT, 3-class	Double attention CNN	Acc. 98.12%; CV $97.86 \pm 0.19\%$	Patient-level split; attention maps
MiniConvNet [3]	CT, NSCLC	Lightweight CNN	Test acc. 96%; 5-fold $\approx 96.6\%$	Compact and fast
LungDxFormer [4]	CT nodules	CNN + transformer	Acc. 97.35%	Repeated runs; $p < 0.01$ vs baselines
Capsule model [5]	CT, 3-class	EfficientNet + capsule	Holdout 100%; 10-fold 92.61%	0.01 GFLOPs; strong split sensitivity
3D Trans-DenseUNet++ [6]	CT nodules	3D segmentation + classifier	Acc. 94.9%	Volumetric localization
DbMCA [7]	CT images + masks	Dual branch	Acc. 98.04% (5k experiment)	Image-mask complementary features
30-model benchmark [8]	CT, 3-class	CNN/ViT benchmark	Top F1 0.9816	Architecture-family comparison
ARXAF-Net [9]	CT, binary	RL + attention + XAI	Test acc. 99%	Active labeling and Grad-CAM
Stacked fusion [11]	CT + clinical + serum	Radiomics + DL + stacking	Test AUC 0.823	Multimodal risk stratification

<i>Study</i>	<i>Task / modality</i>	<i>Method family</i>	<i>Results</i>	<i>Main comparative value</i>
GGN fusion [12]	CT GGN	Radiomics + ResNet	Fusion AUC 0.865	SHAP and DeLong testing
Radiomics vs readers [13]	CT nodules	SVM radiomics	AUC 0.941; acc. 88.6%	Comparable to radiologist consensus
Reader-assistance AI [14]	CT, early cancer	DL + nodule indexing	Reader AUC 0.652→0.761	Sensitivity gain; 12.9% faster reading
AI Brock score [15]	Screening CT	AI risk score	AUC 0.72 vs Brock 0.76	Higher sensitivity, lower specificity
Biology-aware radiomics [16]	CT indeterminate nodules	Radiomics + harmonization	AUC 0.74	Acquisition/biology harmonization
AITO-PETCT-MP [17]	PET/CT nodules	DL malignancy score	AUC 0.78 vs Herder 0.73	Comparable to expert readers
PSCABTM [18]	CT, LUNA16	Cross-attention ensemble	Acc./F1 99.20%	Fine-grained + long-range features
VGG16-U-Net++ [19]	CT	Segmentation ensemble	Acc. 96%	3.1 GB reported GPU memory
Abdullah et al. [20]	CT nodule detection	Enhancement + 3D region growth	Sensitivity 0.92; overlap 0.83	1.5 false positives/scan

Source: Compiled from the cited studies (author-reported values)

4. Comparative Results and Discussion

4.1 Accuracy and Computational Burden Form a Trade-off

The 2026 literature weakens the assumption that high accuracy requires large models. MiniConvNet reports 96% CT test accuracy with a deliberately compact architecture [3], and the capsule-inspired model requires only 0.01 GFLOPs after removal of dynamic routing and decoder reconstruction [5]. Such efficiency can matter more than a small accuracy difference for screening workstations, edge deployment or rapid triage. At the other end, transformer and attention hybrids provide richer contextual modelling, especially for subtle nodules, but larger parameter counts and more complex feature interactions can raise deployment cost [1], [4], [8], [18].

Near-perfect hold-out scores also require careful reading. The capsule-inspired model reports 100% accuracy on a selected hold-out set, yet its ten-fold mean accuracy is 92.61% [5]. This does not invalidate the model; rather, the gap shows the practical value of repeated evaluation. Deployment decisions should therefore consider stable cross-validation,

computational demand and task complexity rather than headline accuracy alone.

4.2 Clinical Utility Can Appear Without the Highest AUC

Reader-assistance evidence provides a different benchmark. Schroeder et al. evaluated 16 board-certified radiologists on 340 CT examinations; AI assistance raised cancer-detection AUC from 0.652 to 0.761, increased mean sensitivity from 0.585 to 0.727, preserved specificity near 0.91 and reduced interpretation time by 12.9% [14]. Such evidence demonstrates an operational benefit that a conventional image-classification test set cannot reveal.

Leijten et al. provide another clinically grounded comparison. Their PET/CT deep-learning malignancy model reached an AUC of 0.78, compared with 0.73 for the guideline-based Herder model and approximately 0.80 for experienced clinicians [17]. Although this is below many public-dataset classifiers, the clinical question is harder: the malignancy probability of indeterminate nodules, with direct implications for surveillance, biopsy or treatment. Clinical relevance therefore rises as the evaluation target moves closer to patient management.

4.3 Risk Scores Reveal Sensitivity–Specificity Consequences

Long et al. reported an AUC of 0.72 for an AI-generated Brock score and 0.76 for the standard Brock score, with a non-significant overall difference near the conventional threshold [15]. The AI score showed higher sensitivity but lower specificity at the tested thresholds. Such a pattern can still be clinically valuable when screening priorities emphasise reducing missed cancers, but increased false-positive follow-up becomes an explicit cost. Closed-set accuracy does not expose such management trade-offs.

4.4 Radiomics and Multimodal Fusion Add Structured Evidence

Radiomics remains competitive in 2026. Selvam et al. report an AUC of 0.941 and 88.6% accuracy for a support vector machine (SVM) built from selected radiomic features, with no significant difference in accuracy from expert radiologist consensus [13]. Huchthausen et al. show a different advantage: biology-aware harmonisation combined with augmented training raises the early-nodule test ROC-AUC to 0.74, illustrating the importance of acquisition effects when lesions remain difficult for standard care [16].

Multimodal fusion supplies information missing from image-only networks. Zong et al. combine clinical variables, radiomics, deep-learning probabilities and serum autoantibodies, and their stacked model reaches a test AUC of 0.823 with 80.3% accuracy [11]. He et al. combine CT radiomics with deep features to predict ground-glass-nodule infiltration and report a fusion AUC of 0.865, supported by SHAP interpretation and DeLong testing [12]. Fusion therefore offers a route to risk stratification

when morphology alone cannot represent the full biological state.

4.5 Explainability Requires Task-Relevant Validation

Grad-CAM and attention maps appear frequently in 2026 systems [1], [2], [9], [10]. Visual explanations help identify whether predictions concentrate on pulmonary lesions rather than irrelevant borders or background structures. Heatmaps alone, however, do not establish clinical validity; radiologist review, feature stability, error analysis and outcome-oriented evaluation provide stronger evidence. Radiomics can complement deep explanations through named shape, intensity and texture variables, while multimodal models can expose the contribution of clinical biomarkers and imaging probabilities.

5. Deployment-Oriented Model Selection

A practical selection framework emerges from the comparative evidence (Table 2). High-throughput or hardware-limited settings favour compact CNN or lightweight capsule designs when adequate local validation is available [3], [5], [19]. Complex image patterns and uncertain nodule boundaries favour attention hybrids or segmentation-assisted pipelines [1], [2], [4], [6], [18]. Continuous malignancy-risk estimation favours radiomics, PET/CT deep learning or multimodal probability fusion [11]–[17]. Workflow integration favours systems evaluated with clinicians, because reader-assistance studies reveal sensitivity, specificity, time and management effects that isolated model metrics cannot show [14], [17].

Table 2. Practical Model-Selection Guide

<i>Use case</i>	<i>Preferred family</i>	<i>Priority evidence</i>	<i>2026 examples</i>
Resource-limited deployment	Compact CNN / lightweight capsule	Model size, FLOPs, latency, stable CV	MiniConvNet [3], capsule model [5]
Fine-grained CT classification	CNN–transformer / attention hybrid	F1, AUC, repeated evaluation	C-Swin [1], LungDxFormer [4], PSCABTM [18]
Nodule boundary dependence	Segmentation-assisted pipeline	Dice/IoU + downstream classification	3D Trans-DenseUNet++ [6], Abdullah et al. [20]
Indeterminate nodule risk	Radiomics / PET-CT risk model	AUC, calibration, threshold behavior	Selvam [13], Huchthausen [16], Leijten [17]

<i>Use case</i>	<i>Preferred family</i>	<i>Priority evidence</i>	<i>2026 examples</i>
Multisource risk stratification	Probability stacking / fusion	AUC, DCA, interpretability	Zong [11], He [12]
Clinical workflow support	Reader-assistance AI	Reader AUC, sensitivity, specificity, time	Schroeder [14]

Source: Compiled by the authors from the cited studies

6. Limitations

The comparative evidence is substantially heterogeneous. CT datasets differ in patient count, slice count, nodule prevalence, histology, augmentation and train–test design, and PET/CT, segmentation, multiclass classification and continuous risk prediction represent different clinical endpoints; cross-paper statistics are therefore inappropriate. Several 2026 studies use public datasets with limited demographic or scanner diversity, reader studies and multi-centre evaluations remain fewer than architecture-focused experiments, and efficiency reporting is inconsistent, with FLOPs, memory, latency and hardware often not reported together.

7. Conclusion

Lung-cancer AI research published in 2026 demonstrates strong technical performance across lightweight CNNs, transformer hybrids, capsule-inspired models, segmentation pipelines, radiomics and multimodal fusion. The most useful comparison no longer concerns the single highest accuracy: model suitability depends on the diagnostic task, computational budget, explanation requirement, need for risk calibration and degree of clinical integration. Compact models provide credible options for constrained deployment; attention hybrids support complex image representation; radiomics and fusion support structured malignancy-risk estimation; and reader-validated AI offers the clearest evidence of workflow impact. Future comparative studies should report performance, calibration, computational cost, explanation quality and reader or guideline comparison together, which would make model selection more clinically meaningful than accuracy-first ranking.

References

- [1]. Yousafzai, S. N., et al. (2026). A hybrid deep learning approach integrating CNN and transformer for lung cancer classification using CT scans. *Scientific Reports*, 16, Article 15420. <https://doi.org/10.1038/s41598-026-41161-7>
- [2]. Aaseegha, M. D., & Venkataramana, B. (2026). Enhancing lung cancer classification through a double attention hybrid CNN-HiFuse approach. *Scientific Reports*, 16, Article 13099. <https://doi.org/10.1038/s41598-026-42290-9>
- [3]. Baqir, M. A., et al. (2026). A lightweight CNN for enhanced non-small cell lung cancer classification using CT scan image. *Scientific Reports*, 16, Article 12985. <https://doi.org/10.1038/s41598-026-41401-w>
- [4]. Srinivasa Rao, K., et al. (2026). LungDxFormer: A transformer-CNN hybrid model with dynamic spatial attention for accurate lung cancer detection and classification. *Scientific Reports*, 16, Article 19233. <https://doi.org/10.1038/s41598-026-50449-7>
- [5]. Bhattacharee, A., et al. (2026). Multiclass lung cancer detection using a hybrid capsule inspired deep neural network. *Scientific Reports*, 16, Article 18615. <https://doi.org/10.1038/s41598-026-49378-2>
- [6]. Illa, P. K., & Thillaigovindan, S. K. (2026). An intelligent lung nodule classification model using 3D Trans-DenseUnet++-based lung nodule segmentation. *Scientific Reports*, 16, Article 19778. <https://doi.org/10.1038/s41598-026-49895-0>
- [7]. Shweikeh, E., et al. (2026). A deep learning model to enhance lung cancer detection using “dual-branch” model classification approach. *PLOS ONE*, 21(1), e0339404. <https://doi.org/10.1371/journal.pone.0339404>

- [8]. Yilmaz, M. T., Algul, E., & Pacal, I. (2026). Robust artificial intelligence frameworks for lung cancer subtyping and malignancy detection on thoracic CT. *Scientific Reports*, 16, Article 27223. <https://doi.org/10.1038/s41598-026-57783-w>
- [9]. Nady, G., et al. (2026). Explainable active reinforcement deep learning improves lung cancer detection from CT images. *Scientific Reports*, 16, Article 7510. <https://doi.org/10.1038/s41598-026-38239-7>
- [10]. Di Giammarco, M., et al. (2026). A method for lung cancer detection by means of explainable convolutional neural networks. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-62967-5>
- [11]. Zong, J., et al. (2026). Stacked CT radiomics, deep learning and clinical feature models for differentiating benign and malignant solitary pulmonary nodules. *Scientific Reports*, 16, Article 19680. <https://doi.org/10.1038/s41598-026-49720-8>
- [12]. He, L., et al. (2026). Prediction of infiltration degree of ground-glass nodules using a fusion of CT radiomics and deep learning. *Scientific Reports*, 16, Article 19153. <https://doi.org/10.1038/s41598-026-50328-1>
- [13]. Selvam, M., et al. (2026). Radiomics-derived classifier performance evaluation in lung nodule characterization compared with expert radiologists. *Scientific Reports*, 16, Article 16629. <https://doi.org/10.1038/s41598-026-47656-7>
- [14]. Schroeder, J. L., et al. (2026). Deep learning model with nodule indexing tailored to early-stage lung cancer detection. *Journal of the American College of Radiology*, 23(6), 984–995. <https://doi.org/10.1016/j.jacr.2026.01.025>
- [15]. Long, K. J., et al. (2026). Evaluation of an artificial intelligence generated Brock score to determine malignancy risk of screen-detected pulmonary nodules. *European Journal of Radiology Open*, 16, Article 100765. <https://doi.org/10.1016/j.ejro.2026.100765>
- [16]. Huchthausen, C., et al. (2026). Training set augmentation and biology-aware harmonization improve radiomic models for lung cancer prediction in indeterminate nodules. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-57670-4>
- [17]. Leijten, L., et al. (2026). Deep learning-based malignancy probability estimation of pulmonary nodules in PET/CT imaging. *European Radiology*. <https://doi.org/10.1007/s00330-026-12744-9>
- [18]. Lakshmi, P., et al. (2026). Parallel spatial cross attention ensembled deep learning model for lung cancer diagnosis and classification using computed tomography images. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-66729-1>
- [19]. Sanjeevkumar, B., et al. (2026). Lung cancer detection in CT scans using extracted nodule features and precise multistage segmentation with deep learning based ensemble model. *Scientific Reports*, 16, Article 25983. <https://doi.org/10.1038/s41598-026-50656-2>
- [20]. Abdullah, R. Y., Venkatesan, C., Naresh, E., & Pradeep Kumar, B. P. (2026). AI driven hybrid convolutional and transformer based deep learning architecture for precise lung nodule classification. *Scientific Reports*, 16, Article 3257. <https://doi.org/10.1038/s41598-025-34569-0>